

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
23 August 2001 (23.08.2001)

PCT

(10) International Publication Number
WO 01/61987 A2

(51) International Patent Classification⁷: H04N
(21) International Application Number: PCT/US01/40013
(22) International Filing Date: 1 February 2001 (01.02.2001)
(25) Filing Language: English
(26) Publication Language: English
(30) Priority Data:
09/505,080 16 February 2000 (16.02.2000) US
(71) Applicant (for all designated States except US): **VERANCE CORPORATION** [US/US]; Suite 500, 6256 Greenwich Drive, San Diego, CA 92122 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **PETROVIC, Rade** [US/US]; 11416 Cypress Woods Drive, San Diego, CA 92131 (US). **WINOGRAD, Joseph, M.** [US/US]; 49 Church Hill Avenue, Cambridge, MA 02140 (US).

(74) Agent: **LIPSITZ, Barry, R.**; Law offices of Barry R. Lipsitz, 755 Main Street, Building No. 8, Monroe, CT 06468 (US).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

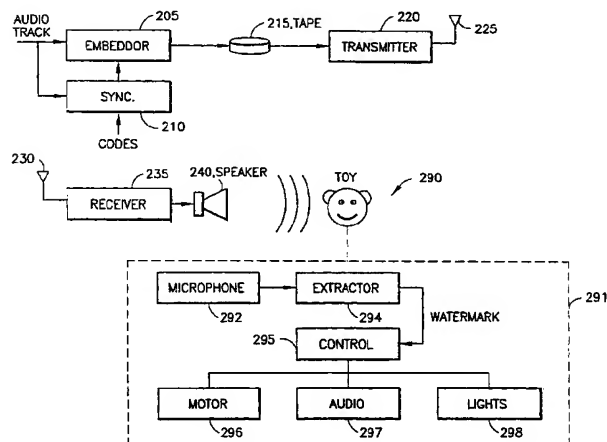
(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: REMOTE CONTROL SIGNALING USING AUDIO WATERMARKS



(57) Abstract: A system for using a watermark embedded in an audio signal to remotely control a device (290). Various devices such as toys, computers, and appliances, equipped with an appropriate detector (292), detect the hidden signals, which can trigger an action, or change a state of the device. The watermarks can be used with a "time gate" device, where detection of the watermark opens a time interval within which a user is allowed to perform an action, such as pressing a button, typing in an answer, turning a key in a lock, etc. To prevent fraudulent activation of a time gate, the time gate device can be configured to react only to watermarks coming from live broadcasts, and not from replays from tapes or other storage devices. In another feature, robustness of the watermark is improved, e.g., for acoustic propagation channels, by shifting the detection time of the watermark based on a measured bit error count of the watermark. Furthermore, the watermark may be inserted before the desired action along with corresponding offset information if the audio signal is not suitable at the time of the action.

REMOTE CONTROL SIGNALING USING AUDIO WATERMARKS

BACKGROUND OF THE INVENTION

The present invention relates to a method and apparatus for remotely controlling a device, such as a toy, lock, smart card, or home appliance, via a control message that is imperceptibly embedded in an audio signal, e.g., as a "watermark". Moreover, the invention optionally enables the device to be synchronized with the audio signal, for example, so that the actions of a doll can be synchronized with a children's television program.

Audio signals are ubiquitous, being broadcast over AM/FM radio, TV, public announcement systems, transmitted over telephone channels, or stored on cassette tapes, CDs, computer memories, etc. Therefore, it is convenient to use audio channels or audio storage to transmit or store some other information.

Audio watermarking, or embedded signaling, has recently emerged as a technology for embedding auxiliary data imperceptibly in a host audio signal. A basic feature of audio watermarking techniques is that the embedded signal is substantially imperceptible to a listener of the host signal. Furthermore, the audio-watermarks occupy the same time/frequency/space domain as the host signal, so that they are not lost in standard audio signal processing, recording or

transmissions, nor filtering and/or masking operations in a deliberate attack can remove them.

5 A primary proposed use of watermarking is in protecting intellectual property rights, e.g., through copy control, automatic broadcast monitoring, ownership dispute resolution, Internet commerce tracking, etc. Alternative applications include auxiliary data embedding, such as the song title and purchasing instructions, assurance of content integrity, proof of performance in TV and radio advertisements, audience exposure monitoring, caller identification (authentication) in telephone conversations, or generic covert communication.

15 Moreover, various schemes have been proposed for sending command and control signals, or their equivalent, concurrently with audio signals. However, these schemes do not qualify as audio watermarking techniques. For example, in one proposed scheme, an "instructional signal" is inserted in a narrow frequency band set aside at the upper frequency edge of the audio spectrum. However, this system does not qualify as a watermarking system since the host and the control signals occupy distinct frequency bands.

25 In another proposed scheme, a unique code describing an offer for products and services is transmitted by a TV program as an audible "beep". There is no attempt to hide this beep, so this technique also is not audio watermarking.

30 In yet another proposed scheme, information related to a TV game show is encoded in touch tones and

broadcast in-band with an audio portion of the show. The touch-tones can be masked by the show's usual sound effects, such as buzzers and beeps. This is substantially different from the watermarking approach, because it cannot simultaneously meet the inaudibility requirement and the requirement for the time domain overlap of a watermark and an arbitrary audio signal.

Accordingly, it would be desirable to provide a watermarking system for sending command and control signals concurrently with audio signals that overcomes the disadvantages of the existing proposed schemes.

The system should use watermarking techniques to provide a hidden data channel in an audio signal for providing short messages, such as device activation commands, or remote control signals that can change the state of a device.

The system should be compatible with existing watermarking techniques, such as those disclosed in U.S. Patent 5,940,135 to Petrovic et al., entitled "Apparatus and Method for Encoding and Decoding Information in Analog Signals," issued August 17, 1999, and incorporated herein by reference.

The system should provide a hidden remote control signal as a watermark within an audio signal for controlling various devices that detect the hidden signal.

The system should allow the remote control signal to be related to, or independent of, content of the host audio signal. For related content, the system

should optionally provide synchronization of the remote control signal with the host audio signal content.

5 The system should use a watermark to define a time gate (window) during which a device is enabled to receive a user input or perform a specified action.

 The system should provide a security mechanism to ensure that the time gate is defined only from a real-time broadcast audio signal, and not from a replay of the audio signal.

10 The system should improve the robustness and temporal resolution of a watermark, and provide a simplified watermark detector.

 The system should provide synchronization of a watermark encoder and decoder.

15 The present invention provides a system having the above and other advantages.

SUMMARY OF THE INVENTION

The present invention relates to a system for using a watermark embedded in an audio signal to remotely control a device.

5 In particular, the system is compatible with existing audio-watermarking technologies that use audio channels and/or audio storage to carry independent data without interfering with the audio channel's original purpose. However, such a channel has much lower
10 information capacity than a modem channel, typically no more than about twenty bits per second per audio channel. The invention uses this hidden data channel for relatively short messages, such as device activation commands, or remote control signals that can
15 change state of a device.

A remote control signal is hidden within an audio signal that is broadcast over radio and TV, stored on CDs, DVD, tape or computer memory, played over speakers and/or transmitted over other audio channels. Various
20 devices such as toys, computers, and appliances, that are equipped with an appropriate detector, detect the hidden signal to trigger an action, or change a state of the device. The device action can be completely unrelated to the ongoing audio content, and it can have
25 a number of different objectives, such as entertainment, education, sales, security, etc.

In one particular implementation, a "time gate" device is disclosed, where detection of the watermark opens a time interval within which a user is allowed to

perform an action, such as pressing a button, typing in an answer, turning a key in a lock, etc. To prevent fraudulent activation of a time gate, the time gate device can be further upgraded to react only to
5 watermarks coming from live broadcasts, and not from replays from tapes or other storage devices.

In another implementation, detection of the watermark triggers an action.

Additionally, techniques are presented for
10 improving existing watermarking technology in view of requirements for the proposed applications. In particular, the invention provides improvements in robustness of the watermark in the channels with acoustic propagation (e.g., propagation through air) -
15 using delay hopping (watermarking adjacent bits using distinct autocorrelation delays), robustness improvements using redundant watermarking, improvements in the time resolution of the trigger feature, and simplifications of the detector design.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 illustrates an audio watermarking process in accordance with the present invention.

5 FIG. 2 illustrates a system for remote control of a device, such as a toy, in synchronism with audio data, such as from a television program, in accordance with the present invention.

10 FIG. 3(a) illustrates a time gate defined by start and stop watermarks in accordance with the present invention.

FIG. 3(b) illustrates a time gate defined by a start watermark and a fixed interval τ in accordance with the present invention.

15 FIG. 3(c) illustrates a time gate defined by a start watermark and a multiple N of a fixed interval τ' in accordance with the present invention.

FIG. 4 illustrates a real-time time gate application in accordance with the present invention.

20 FIG. 5 illustrates the use of countdown watermarks for defining a start time of a desired action in accordance with the present invention.

FIG. 6 illustrates an autocorrelation modulation extractor based on sign correlation in accordance with the present invention.

25 FIG. 7 illustrates an example of a bit error count versus a time shift for detecting a watermark in accordance with the present invention.

DETAILED DESCRIPTION OF THE INVENTION

The present invention relates to a system for using a watermark embedded in an audio signal to remotely control a device.

5 FIG. 1 illustrates an audio watermarking process in accordance with the present invention.

At an encoding side 100, a watermark, i.e., an embedded signal, is inserted into an audio signal at an embeddor 105, using a key, which is a set of parameters that define the hiding process. The key may comprise a steganographic key. The composite signal that is output from the embeddor 105 can be recorded, transmitted over various channels, or processed in different ways, which usually includes corruption by noise and distortion.

10

15

The composite signal is received at a decoding side 150, where the embedded signal (watermark) is retrieved from the composite signal in an extractor 155 with the help of the key used in the embedding process.

20 Various details regarding conventional signal processing techniques, such as compression, coding, error-correction, modulation, and the like, are not explicitly disclosed but their appropriate use will be evident to those skilled in the art.

25 The embeddor 105 may provide the watermark in the audio signal using various known watermarking techniques, including those discussed in the aforementioned U.S. Patent 5,940,135, where the short-term autocorrelation function of the host audio signal

is modulated in such a way to match the embedded signal. The key contains the information about the frequency band of the host signal used for hiding, the delay set used for autocorrelation calculation and its change patterns, baseband symbol waveform, data packet structure, scrambling key, etc. With this system, the extractor 155 calculates the short-term autocorrelation using the same key, and regenerates the inserted message with the help of standard digital signaling techniques.

Typical bit rates of the embedded messages are low, ranging from a few bits per second to a few tens of bits per second. For example, the International Federation of the Phonograph Industry required 20 bps per channel for its review of watermarking technologies. Those skilled in the art can appreciate that the increase in the bit rate brings reduced robustness, increased audibility and/or increased complexity. Therefore, only relatively short messages can be hidden within an audio signal. This is quite suitable for intellectual property protection, where copy control information and/or content identification codes of less than one-hundred bits are embedded.

However, the present invention proposes the use of audio watermarks for remote control of various devices, such as toys, locks, smart cards, appliances, etc., over standard audio channels, such as radio and TV broadcasts, audio tapes, CDs, telephone channels, public address systems, etc. As an example, we will

describe a system for remote control of toys participating in a TV show, as illustrated on FIG. 2.

FIG. 2 illustrates a system for remote control of a device 290, such as a toy, in synchronism with audio data, such as from a children's television program, in accordance with the present invention.

The control messages are inserted using audio watermarking at appropriate places of the show's audio track using an embeddor 205 and a synchronization device 210. The composite audio signal is optionally stored, e.g., on a tape 215, and subsequently broadcast via a transmitter 220 and antenna 225. While antennas 225 and 230 are shown as an illustration, any type of broadcast scheme can be used, including delivery via a terrestrial path, cable, optical fiber, and/or computer network. Moreover, broadcast of the composite signal to a large population of receivers is not required, as the invention is also suitable for any transmission, including point-to-point transmissions, video-on-demand transmissions in a cable television network, and so forth. Moreover, the composite audio signal may be played back locally at a user's home from a storage device such as a tape or disc.

A TV receiver 235 receives the signal via an antenna 230 and the audio signal is played over a TV speaker 240. The toy 290, specially designed in accordance with the present invention, includes components 291, including a built-in microphone 292 for picking up (detecting) the audio signal, a watermark extractor 294 for extracting the watermarks from the

composite audio signal, and a control 295 that is responsive to the watermarks for performing some action.

5 The toy 290 may optionally be hard-wired to the receiver using appropriate jacks and wiring, in which case the microphone 292 is not needed.

10 A motor 296, audio function 297, such as a speech synthesizer, and lights function 298 are responsive to the control 295. For example, if the toy 290 is a doll, the audio function 297 may play a prerecorded message in concert with the ongoing show. The motor 296 may cause the toy's 290 arms, legs, head and/or mouth to move. The lights function 298 may cause the toy's eyes to light up. This creates the appealing
15 impression that the toy is actually following the show together with the children, and that it participates.

20 Advantageously, no modification is required for the TV channel, including the TV signal storage equipment, satellite distribution channels, broadcast equipment, and TV sets. Additional equipment consists only of an embeddor, specially designed for precise watermark insertion at a desired segment of the audio track, and a mass-produced, inexpensive detector incorporated in a suitably-designed toy and connected
25 to the toy controller 295. Also, note that the same toy 290 can be activated by audio watermarks coming from any audio source, such as an AM/FM radio broadcast, CD or tape player, or speakers wired to a computer.

An important feature in the previous example is the synchronization of the action of a toy (or toys) with the ongoing show. To achieve this, the watermark should be embedded in the audio track segment immediately preceding the desired moment of the toy action, with a small allowance for processing and propagation delays. This synchronizing feature can be useful in many other practical applications.

Time Gate

For example, we will describe here a device suitable for watermark-based activation that we will call a "time gate" (e.g., window). In accordance with the invention, the detection of a watermark causes the time gate to open a time interval, during which a user is allowed to perform an action. For example, during an interactive TV quiz show, the viewers may participate by keying in their answers to a hand-held unit, while the players in the television studio prepare their answers. This can be achieved if the audio track has appropriately inserted watermarks, and if a time gate device synchronized to the watermarks controls access to the hand-held unit.

Such a hand-held unit can have similar componentry as the toy 290 of FIG. 2. The control 295 can be configured to send a message to the user to alert the user that the time gate has started or ended, the duration of the time gate, and the amount of time remaining in the time gate. An output screen on the hand-held unit, such as a liquid-crystal display (LCD), may be appropriate for this purpose. Or, the hand-held

unit may send a signal back to the television receiver 235 to provide a display on a TV or computer screen that informs the user of the information provided by the watermark (e.g., via a wired path, infra-red signal or the like).

Alternative applications including an automated audio/video exam, where the time gate defines a period during which the user can enter responses for the exam, alertness monitoring, where the user is required to provide an input during the time gate, TV coupon collection, where electronic coupons can be retrieved by a user during the time gate, remote control of a lock, where the lock can be opened or closed only during the time gate, and so forth, are discussed below.

Three different designs of the time gate protocol are illustrated in FIGs 3(a)-3(c).

FIG. 3(a) illustrates a time gate defined by start and stop watermarks in accordance with the present invention.

This design uses two distinct watermarks, including a start watermark 305 and a stop watermark 310, to mark the desired beginning and the end, respectively, of the gated interval 315 bounded by T_1 and T_2 . Note that the boundaries of the interval 315 are shown as occurring slightly after the end of the corresponding watermarks, due to processing and propagation delays. This design can mark an arbitrary time interval that is larger than the duration of the

stop watermark 310, since the duration of the stop watermark 310 consumes a portion of the time gate 315.

FIG. 3(b) illustrates a time gate defined by a start watermark and a fixed interval τ in accordance with the present invention.

This time gate protocol design is simpler than that of FIG. 3(a) since it requires a single start watermark 330 before the beginning of the marked interval 335. The duration of the marked interval (time gate), τ , is predefined, e.g., at the time the device, such as the toy 290 in FIG. 2, is manufactured. However, the interval may be re-programmable (e.g., by replacing a memory chip). This design is suitable if the user action is simple, like pressing a button, or turning a key in a lock.

FIG. 3(c) illustrates a time gate defined by a start watermark and a multiple N of a fixed interval τ' in accordance with the present invention.

This time gate design requires a single watermark 350, but still can mark a variable time interval. This is achieved by inserting an integer number N into the watermark's data field, where the interval 360 is N times a predefined time slot duration τ' .

The value N can be carried in a separate or same watermark as that of the control data, or even provided beforehand. A preferred solution is for the same watermark 350 to include both N and the control data (for reasons of efficiency, potential for false action, etc.)

FIG. 4 illustrates a real-time time gate application in accordance with the present invention.

Many time gate applications allow pre-processing of the audio track in a similar manner as depicted in FIG. 1, e.g., for toy applications. For example, the time gate device can be used in an educational process to monitor alertness of the student. In this case, a pre-recorded lecture (video/audio, or audio only) contains a watermarked content that opens a time gate interval after questions/instructions that require the student's response, such as pressing yes/no buttons on the time gate device. The device may score correct/incorrect answers for review by a teacher, but we expect major educational benefits by simply confirming that the student was paying attention to the audio signal.

Similarly, an advertisement company may poll viewers about an advertised product, and engage their attention. Active participation of users is expected to bring better recognition of products in a shorter time, and lower advertisement costs. In particular, as an encouragement for participation, an advertiser may offer discounts to those who bring the time gate device to a specified retail store, where the user's answers to the poll questions are downloading from the device. This is equivalent to bringing cut-out coupons from a paper advertisement. Again, the watermarks are pre-stored in the audio track of a TV or radio adds.

However, in other applications, the embedding of the watermark should occur in real-time, while the

audio signal is being transmitted. Such a case occurs when the time gate is used, for instance, to control a user's access to a secure area or safe box, or to override other restrictions. For example, an innocuous public address system announcement or background music may be used with an electronic lock to allow a key to operate within the defined time interval. Similarly, a user may remotely control a door lock in the user's residence by calling home over a telephone line, and speaking to the telephone answering machine (with its speaker on). The user's telephone set includes an encoder that embeds watermarks with control data into the user's voice. When the voice is received at a decoder with a watermark extractor at an electronic door lock mechanism, the lock can be activated.

In a related example, background music or a public address announcement in a larger facility such as an office may be used as a host audio signal for watermarked control data to remotely lock or unlock doors, filing cabinets and so forth without requiring re-wiring of the facility.

With these application, it is important that there is no detectable communication channel between the user and the controlled device, so that potential attackers are oblivious to the control mechanism. Furthermore, if an attacker does learn the operational principle of the lock, he/she should not be able to gain access to the secure area by recording and subsequently replaying the audio signal with the embedded control data.

Even if an attacker knows the operational principle of the lock, he/she cannot forge the message if he/she does not know the secret key used in the embedding process, as shown in FIG. 1.

5 The real-time time gate system shown in FIG. 4 addresses these concerns. An encoding side 400 includes an audio source 405, an embeddor 410, a clock 415 and a control data encoder 420. A receiving side 470 that receives the real-time transmission of the
10 composite audio signal includes an extractor 475, a logic function 480, and a clock 485.

 The embeddor 410 receives a continuous audio signal from the audio source 405, and also receives a string of messages from the data encoder 420. The
15 messages include the control data, as well as timing information from the local clock 415, with possibly some additional data (e.g., number of slots for slotted time gate). The message embedding is triggered from outside, e.g., by an operator pressing a button on the
20 embeddor 410. The message is inserted into the audio data stream using a secret key. In the real-time operation, the message is immediately transmitted, i.e., there is no recording of the audio signal except for a short buffering in the embeddor 410 necessary for
25 the embedding process (up to a few tens of milliseconds). The output composite signal is transmitted over a standard audio channel (telephone, radio, TV, public announcement system, etc). The receiver 470 detects the signal and passes it to the

extractor 475, which detects the message and passes it to the logic function 480 for verification.

5 The logic function 480 compares the timing data from the incoming message to locally-generated timing data from the clock 485. If the match is sufficiently close, the logic function 480 concludes that a real-time transmission has occurred, and the control information taken from the received message is passed, e.g., starting the time gate. Simultaneously, the
10 local clock 485 may be adjusted (e.g., synchronized with the clock 415) if the discrepancy with the transmitter clock 415 is within predefined bounds. However, if the time discrepancy is too large (e.g., the local time is significantly after the transmitted
15 time), it implies that recording and play back may be taking place, so the message is ignored.

 If the local time is significantly before the transmitted time, this implies a significant mis-calibration, and a default mode can be invoked to
20 ignore the message.

 Clearly, potential attacks based on storing and replaying messages would introduce large delays, certainly more than a few seconds. On the other hand, propagation and processing delays are certainly less
25 than a second, so there should be a clear separation of the two cases. Also, typical clock devices can maintain a timing error well below a couple of seconds for a quite long time, so that the time drift of local clocks can be also distinguished from the store-and-replay attack. Occasional re-synchronization of the
30

receiver clock with the transmitter's clock should keep the timing mismatch between the send and receive clocks within tight bounds indefinitely. This re-synchronization can be achieved using known techniques.

5 Watermark design considerations

The applications described above impose somewhat different sets of requirements on the audio watermark design than do the intellectual property protection applications. For example, most of the applications described above imply an acoustic (free-space) propagation channel, while the property rights establishment is usually done on an electronic form of the audio signal. The use of an acoustic propagation channel is especially challenging since it generates intersymbol interference due to multipath, signals may be corrupted with background noise, and the distortions in speakers and microphones are usually large compared to electronic channel distortions.

Furthermore, some of the above applications, such as the toy applications, require very inexpensive designs for the watermark extractors. On the other hand, in some other applications, such as remote control of locks, it is equally important to have an inexpensive embeddor as well, e.g., for the telephone set example. Finally, the time gate device requires careful consideration of timing tolerances to achieve the best possible resolution in time domain, as opposed to typical copyright protection applications, where the location of the watermark within a signal is relatively unimportant.

With the above requirements in mind, the present inventors looked at different watermarking technologies, and found that the best-suited technology is the autocorrelation modulation (ACM) technique described in the aforementioned U.S. Patent 5,940,135. However, any other suitable watermarking/data embedding technique may be used.

ACM features a simple design for the embeddor and extractor, high robustness, large throughput, low probability of falses, good layering capability, and full inaudibility. However, the present invention provides improvements and the special selection of design parameters to optimize ACM for the proposed applications. Herein, several techniques are disclosed that can substantially improve the performance of ACM in remote control signaling using audio watermarks.

The acoustic-coupling environment, where a speaker broadcasts an audio signal, and the detector captures it through a microphone, can be improved by a special watermark design that is not addressed by conventional techniques. The main issue is the multipath propagation caused by reflections of acoustic waves, which may introduce intersymbol interference to the watermark detector. Standard techniques to fight intersymbol interference, such as adaptive equalization, are too costly for an inexpensive detector. An increase in the bit interval is helpful, but with obvious drawback in the reduction of the watermark bit rate.

To avoid intersymbol interference, we propose watermarking adjacent bits using distinct autocorrelation delays ("delay hopping"). In effect, distinct autocorrelation delays can be considered as
5 distinct channels, with little interference between them. This aspect of the invention increases the watermark robustness, which is particularly useful in the acoustic-coupling environment, but may be used in other environments as well. That is, if consecutive
10 symbols are sent over distinct channels, they cannot cause intersymbol interference, regardless of the pulse broadening caused by a multipath environment.

In a further aspect of the invention, robustness of the watermarking is improved by first evaluating the
15 masking ability of the host audio signal before embedding the watermark. In a typical scenario, the device activation, time gate opening, or other actions occur upon detection of the corresponding watermark. This means that the encoder should insert the watermark
20 immediately prior to the desired moment, taking into account propagation and processing delays. However, the desired (candidate) watermark insertion interval may be unsuitable, for example, if it is mainly a silence. Accordingly, the watermark can be inserted
25 before the desired instant of action, along with information for informing the decoder about the delay between the watermark detection and the desired action, which corresponds to the delay between the time segment in which the watermark is embedded and the original
30 desired time segment.

There is a tradeoff between the flexibility in choosing the optimum watermark insertion time and the amount of the payload (bits) assigned to the delay information.

5 FIG. 5 illustrates the use of countdown watermarks for defining a start time of a desired action in accordance with the present invention.

 In a further alternative embodiment, a string of watermarks 510 is embedded before the desired start of
10 action (T_1). Each watermark, such as watermarks 512, 514 and 516, has a countdown data field ($n=2, 1, 0$) that indicates the number of remaining watermarks before (T_1). Detection of any of the watermarks in the string 510 allows the calculation of the desired timing
15 of the action. For example, if the countdown field of a particular watermark contains the countdown field value n , and the watermark duration is w seconds, than the desired action should begin $n*w$ seconds after the particular watermark is detected, plus some additional
20 propagation and processing delay.

 This provides improved robustness since the start time of the desired action is designated with
25 redundance. Thus, even if all but one of the watermarks are not received correctly, the start time will be still be designated.

 FIG. 6 illustrates an autocorrelation modulation extractor based on sign correlation in accordance with the present invention.

 The simple decoder design described in the
30 aforementioned U.S. Patent 5,940,135 includes a filter,

followed by a delay line, correlator (multiplier), and an integrator. The output of the integrator represents a base-band watermark signal without normalization.

When a binary message has been embedded, the output is
5 positive at the decision moment if a "one" bit is embedded, or negative for a "zero" bit. In this design, it is important to maintain a very precise delay in the delay line; small errors in the delay can bring significant distortion of the watermark signal.

10 Those skilled in the art know that precise delays are best achieved by a digital delay line. This means that an A/D converter is necessary prior to the delay line, which add to the cost of the decoder.

In accordance with the present invention, the
15 decoder of U.S. Patent 5,940,135 can be further simplified as shown in FIG. 6 to meet even lower cost demands, suitable for toy applications, without compromising the delay precision.

This decoder 600 includes a filter 610 that
20 receives the composite signal with the watermark, e.g., from a channel or a storage device, and a comparator 620 for comparing the filtered signal to ground 630. An AND gate 640 receives the output of the comparator 620, and a local clock signal. An XNOR (exclusive-NOR)
25 gate 660 receives a direct output of the AND gate 640, as well as a shifted version of the output via shift register 650. The output of the XNOR gate 660 is provided to a counter 670, which communicates with a logic function 680.

Thus, instead of multiplying the received signal with a delayed version of itself, it is possible to detect the signal polarity, and then perform an XNOR operation between the signal and a delayed version of itself (i.e., if the signs are the same, the output is one, if opposite, the output is zero). Then, instead of integration, it is enough to run the counter 670 at a clock rate that is much higher than the bit rate for the duration of the watermark bit. If the count at the end of the bit interval is more than half the maximum count, then a "one" bit is detected; otherwise a "zero" bit is detected. A synchronizing algorithm, residing in the logic block 680, determines the beginning and the end of the bit interval, and generates a reset signal for the counter at the end of each bit.

The above simplification shows that the comparator substitutes for an A/D converter, the XNOR gate 660 replaces a multiplier, and the counter 670 substitutes for an integrator. Moreover, the present inventors have confirmed through experimentation that the proposed simplification does not significantly reduce robustness.

FIG. 7 illustrates an example of a bit error count versus a time shift for detecting a watermark in accordance with the present invention.

In the time gate applications in particular, but in other triggering applications as well, it is important to achieve a good timing precision in detecting the end of a watermark. In the case of watermarks described in U.S. Patent 5,940,135, this is

equivalent to detecting the trailing edge of the last bit of the watermark bit stream. However, noise and channel distortions can corrupt the trailing edge of any bit, or the bit as a whole. Therefore, it is
5 necessary to take the watermark in its totality to decide the most probable timing of the end of the watermark.

In typical digital watermarks described in U.S. Patent 5,940,135, error correction codes are used to
10 recover watermarks in the case when some bits are corrupted. In this case, the watermark can be detected at a time interval that is slightly earlier than its optimal position, but with a higher error count than in its optimal position. The present invention modifies
15 this scheme propose to use this feature to further optimize the watermark timing detection.

In accordance with the present invention, the decoder will attempt to detect the watermark with a starting and ending position that are delayed slightly
20 with respect to the timing position where initial detection of the watermark occurred. The time shifts are very small with respect to the watermark duration, typically of the order of 5% of the bit interval (T_{bit}) (each watermark contains tens or even hundreds of
25 bits). In each of these shifts, the decoder will continue to detect the same watermark and monitor the bit error count over the duration of the watermark. The same watermark is detectable even with slight time
30 shifts, which is the basis of the precise time resolution disclosed herein. The optimum timing is

found where the error count is minimized. For multiple minimums, the optimum timing can be taken at the midpoint.

For example, in FIG. 7, the error count changes as the detecting time is shifted. Initially, with no time shift (i.e., at the originally-detected position), the watermark is detected with four bits in error (point 710). However, as the detecting time is shifted in steps of ts (e.g., $ts = T_{bit}/20$), the error count decreases (points 715 and 720), reaches a minimum for two shifts (points 725 and 730), and then increases (points 735, 740 and 745).

A minimum bit error (points 725 and 730) is reached for two consecutive steps, and we conclude that the best estimate for a reference position of the watermark (which can be, e.g., the ending time of the watermark) is at the mid-point of these two events, i.e., at $(ts_3 + ts_4)/2$. Alternatively, either ts_3 or ts_4 could be selected. The resolution of the extractor's clock will govern the minimum possible time shift.

If only one minimum is found, the optimum position is taken at that time shift position.

To reduce computations, it is possible to terminate the bit error calculations once a minimum has been detected and the bit error count begins to rise (e.g., at point 735). Also, optionally, the bit error calculations can be terminated when a bit error count of zero is first reached.

As an numeric example, assume the watermark message is two seconds long, and $T_{bit} = 0.08$ sec. Let us

say that watermark is first detected in the audio signal starting at time 12.340 sec. and ending at time 14.340 sec. This corresponds to a zero time shift. Then, with $ts1=0.08/20=0.004$ sec., the watermark is
5 next detected starting at time 12.344 sec., and ending at time 14.344 sec., and so forth. Note that the detection interval is 2 sec., and the position shift, and the resolution, is 0.004 sec.

Moreover, for a strong watermark, the error count
10 may reach zero or some other minimum value for several time shifts. In this case, the optimum timing is again set in the middle of the time intervals with the minimum errors.

Experiments show that this technique detects the
15 watermark with the precision of $\pm 20\%$ of the bit interval. For example, for watermarks running at a 50 bit/sec. rate, this corresponds to a time resolution of ± 4 ms. This resolution corresponds to ± 1.4 meters of acoustic propagation delay, so it is adequate
20 for the systems based on acoustic coupling.

Accordingly, it can be seen that the present invention provides a method and apparatus with various advantages, including:

- allows remote control of a device, such as a
25 toy, lock, smart card, or home appliance, via a control message that is imperceptibly embedded in an audio signal as a watermark;
- is compatible with, and builds upon, existing watermarking techniques;

- allows the remote control signal to be synchronized with the audio content, such as to allow a toy to move in conjunction with the audio track of a children's television program;
- 5 • uses a watermark to define a time gate (window) during which a device is enabled to receive a user input;
- provides a security mechanism to ensure that the time gate is defined only from a real-time broadcast audio signal, and not from a replay of the audio signal;
- 10 • provides a simplified watermark detector;
- improves the robustness and temporal resolution of a watermark, e.g., in acoustic propagation channels, by shifting the detecting point based on an error count of the watermark; and
- 15 • improves robustness by evaluating the masking ability of the host audio signal when embedding the watermark, and inserting the watermark in a suitable interval before the desired action, along with information indicating the time shift, if necessary.
- 20

Although the invention has been described in connection with various specific embodiments, those skilled in the art will appreciate that numerous adaptations and modifications may be made thereto without departing from the spirit and scope of the invention as set forth in the claims.

25

What is claimed is:

1. A method for remotely controlling a device, comprising the steps of:

embedding at least a first control message for triggering a desired action of the device as a corresponding first watermark in an audio source signal to provide a composite audio signal;

wherein the first watermark is carried substantially inaudibly in the composite audio signal; and

communicating the composite audio signal to the device; wherein:

the device is adapted to receive the composite audio signal communicated thereto and recover the first control message embedded therein for triggering the desired action in accordance therewith; and

the first watermark and the audio source signal coexist in the composite audio signal in a common time, frequency and space domain.

2. The method of claim 1, wherein said embedding step comprises the step of:

modifying the audio source signal so that data symbols of the first control message are determined according to corresponding values of an autocorrelation function of the composite audio signal;

wherein each autocorrelation value is calculated over a time interval of a corresponding data symbol.

3. The method of claim 1, wherein:
the first control message designates a time for
triggering the desired action.

4. The method of claim 1, wherein:
the first control message designates the desired
action.

5. The method of claim 1, wherein:
the composite audio signal is communicated to the
device, at least in part, by propagating via a speaker.

6. The method of claim 1, wherein:
in said embedding step, the first control message
is embedded in the audio source signal so that the
desired action is triggered in synchronism with a
desired content of the audio source signal.

7. The method of claim 1, wherein:
in said embedding step, the first control message
is embedded in the audio source signal so that the
desired action is triggered in synchronism with a
desired content of a video signal that is associated
with the audio source signal.

8. The method of claim 1, wherein:
in said communicating step, the composite audio
signal is communicated, at least in part, via a
broadcast channel.

9. The method of claim 1, wherein:
the composite audio signal is stored at a storage device and subsequently retrieved from the storage device for said communicating step at a designated time for triggering the desired action.

10. The method of claim 1, wherein:
a time gate for triggering the desired action is defined in accordance with the first control message.

11. The method of claim 10, wherein:
a start of the time gate is defined according to the first control message.

12. The method of claim 10, wherein a start of the time gate is defined according to the first control message, comprising the further step of:
embedding a second control message as a corresponding second watermark following the first watermark in the audio source signal;
wherein an end of the time gate is defined according to the second control message.

13. The method of claim 10, wherein:
a start of the time gate is defined according to the first control message, and an end of the time gate is defined according to a fixed gate interval that follows the start.

14. The method of claim 10, wherein:

the first watermark comprises a countdown data field for defining a start of the time gate according to a multiple of a duration of the first watermark.

15. The method of claim 10, wherein:

a start of the time gate is defined according to the first control message, and an end of the time gate is defined according to a multiple of a fixed gate interval that follows the start.

16. The method of claim 10, comprising the further step of:

embedding at least a second watermark in the audio source signal following said first watermark;

wherein each watermark comprises a countdown data field for defining a start of the time gate according to a multiple of a duration of each watermark.

17. The method of claim 16, wherein:

the countdown data field of only one of the watermarks is required to define the start.

18. The method of claim 1, wherein:

the first control message is embedded in the audio source signal according to a steganographic key to control access to the first control message.

19. The method of claim 18, wherein:

the device is adapted to use the steganographic key in recovering the embedded first control message.

20. The method of claim 1, wherein:
the desired action is to change a state of the device.

21. The method of claim 20, wherein:
the state of the device is changed to control access to the device.

22. A method for enabling a device to perform a desired action via remote control, comprising the steps of:

receiving a composite audio signal communicated to the device; and

recovering at least a first control message embedded in the composite audio signal to perform the desired action; wherein:

the first control message is embedded as a corresponding first watermark in an audio source signal to provide the composite audio signal;

the first watermark is carried substantially inaudibly in the composite audio signal; and

the first watermark and the audio source signal coexist in the composite audio signal in a common time, frequency and space domain.

23. A method for defining, by remote control, a time gate for a device, during which an access state of the device is changed, comprising the steps of:

embedding at least a first control message for changing the access state of the device as a corresponding first watermark in an audio source signal to provide a composite audio signal, said first watermark being carried substantially inaudibly in the composite audio signal; and

communicating the composite audio signal to the device;

wherein the device is adapted to receive the composite audio signal communicated thereto and recover the first control message embedded therein for changing the access state in accordance therewith.

24. The method of claim 23, wherein:

the access state is changed from: (a) not allowing a user input to control the device outside the time gate, to (b) allowing a user input to control the device during the time gate.

25. The method of claim 23, wherein:

the access state is changed from allowing a user input to control the device outside the time gate to not allowing a user input to control the device within the time gate.

26. The method of claim 23, wherein:

the access state is changed to allow a user to actuate a key to control the device during the time gate.

27. The method of claim 23, wherein:
the access state is changed to allow a user to
actuate an electronic interface to control the device
during the time gate.

28. The method of claim 23, wherein:
the composite audio signal is communicated to the
device, at least in part, via propagation from a
speaker.

29. The method of claim 23, wherein a start of
the time gate is defined according to the first control
message, comprising the further step of:
embedding a second control message as a
corresponding second watermark following the first
watermark in the audio source signal;
wherein an end of the time gate is defined
according to the second control message.

30. A method for defining, by remote control, a
time gate at a device during which an access state of
the device is changed, comprising the steps of:
receiving a composite audio signal communicated to
the device; and
recovering at least a first control message
embedded in the composite audio signal for use in
changing the access state of the device; wherein:
the first control message is embedded as a
corresponding first watermark in an audio source signal
to provide the composite audio signal; and

the first watermark is carried substantially inaudibly in the composite audio signal.

31. A method for remotely controlling a device, comprising the steps of:

embedding, at an encoder, at least a first control message for triggering a desired action of the device as a corresponding first watermark in an audio source signal to provide a composite audio signal;

embedding clock data into the audio source signal according to an associated clock of the encoder;

wherein the first watermark and clock data are carried substantially inaudibly in the composite audio signal; and

communicating the composite audio signal to the device;

wherein the device is adapted to receive the composite audio signal communicated thereto, recover the embedded clock data and compare it to associated clock data of the device, and, if a difference between the recovered clock data and the device's associated clock data is within a predefined bounds, recover the first embedded control message for triggering the desired action in accordance therewith.

32. The method of claim 31, comprising the further step of:

synchronizing the device's associated clock data with the encoder's associated clock data if the difference is within the predefined bounds.

33. The method of claim 31, comprising the further step of:

precluding the triggering of the desired action if the difference is not within the predefined bounds.

34. The method of claim 31, wherein:

the predefined bounds accounts for an expected time for communicating the composite audio signal from the encoder to the device.

35. The method of claim 31, wherein:

the clock data associated with the encoder is embedded in the audio source signal according to a steganographic key to control access thereto.

36. The method of claim 35, wherein:

the device is adapted to use the steganographic key in recovering the embedded clock data.

37. The method of claim 31, comprising the further step of:

periodically re-synchronizing the device with the audio source signal.

38. The method of claim 31, wherein:

the composite audio signal is communicated in said communicating step substantially without delay after being provided.

39. A method for enabling a device to perform a desired action via remote control, comprising the steps of:

(a) receiving a composite audio signal communicated to the device; wherein:

at an encoder, at least a first control message for triggering a desired action of the device is embedded as a corresponding first watermark in an audio source signal to provide the composite audio signal,

clock data is embedded into the audio source signal according to an associated clock of the encoder, and

the first watermark and clock data are carried substantially inaudibly in the composite audio signal; and

(b) recovering the embedded clock data and comparing it to associated clock data of the device, and, if a difference between the recovered clock data and the device's associated clock data is within a predefined bounds, recovering the first embedded control message for triggering the desired action in accordance therewith.

40. A method for robustly embedding data in an audio signal, comprising the step of:

embedding at least a first message as a corresponding first watermark in an audio source signal to provide a composite audio signal by modifying the audio source signal so that data symbols of the first message are determined according to corresponding

values of an autocorrelation function of the composite audio signal; wherein:

the values of the autocorrelation function are determined using delays that differ for adjacent data symbols; and

the first watermark is carried substantially inaudibly in the composite audio signal.

41. The method of claim 40, wherein:

each autocorrelation value is calculated over a time interval of a corresponding data symbol.

42. A method for recovering data that is robustly embedded in an audio signal, comprising the steps of:

(a) receiving a composite audio signal; wherein:

at an encoder, at least a first message is embedded as a corresponding first watermark in an audio source signal to provide the composite audio signal by modifying the audio source signal so that data symbols of the first message are determined according to corresponding values of an autocorrelation function of the composite audio signal;

the values of the autocorrelation function are determined using delays that differ for adjacent data symbols; and

the first watermark is carried substantially inaudibly in the composite audio signal; and

(b) recovering the first message from the composite audio signal.

43. The method of claim 42, wherein:
each autocorrelation value is calculated over a
time interval of a corresponding data symbol.

44. A method for robustly embedding data in an
audio signal, comprising the steps of:
determining a first, candidate time segment of an
audio source signal for embedding at least a first
message as a corresponding first watermark to provide a
composite audio signal;
evaluating the audio source signal to determine if
the first time segment is suitable for masking the
first message; and
if the first time segment is found to be
unsuitable:
(a) selecting another, second time segment of the
audio source signal that is suitable for masking the
first message; and
(b) embedding the first message in the second time
segment along with information indicative of a time
shift between the first and second time segments.

45. The method of claim 44, wherein:
the second time segment precedes the first time
segment.

46. A method for recovering embedded data from a
composite audio signal, comprising the steps of:
(a) receiving the composite audio signal; wherein:

at least a first message is embedded as a corresponding first watermark in a first time segment of the composite audio signal, along with time shift information indicative of a time shift between the first time segment and another, second time segment; and

an audio source signal of the composite audio signal is suitable for masking the first message during the first time segment, but is unsuitable for masking the first message during the second time segment;

(b) recovering the first message and the time shift information from the first time segment of the composite audio signal;

(c) determining when the time shift has elapsed in accordance with the time shift information; and

(d) providing a control signal for controlling a device after the time shift has elapsed in accordance with said determining step.

47. The method of claim 46, wherein:
the second time segment precedes the first time segment.

48. A method for decoding embedded data in an audio signal, comprising the step of:

(a) receiving a composite audio signal;

wherein data symbols of at least a first message are embedded as a corresponding first watermark in an audio source signal to provide the composite audio

signal, and the first watermark is carried substantially inaudibly in the composite audio signal;

the audio source signal is modified so that data symbols of the first message are determined according to corresponding values of an autocorrelation function of the composite audio signal; and

each autocorrelation value is calculated over a time interval of a corresponding data symbol; and

(b) recovering the embedded data symbols from the received composite audio signal according to polarities of the composite audio signal.

49. A decoding method for improving the time resolution of the position of a message that is embedded in an audio signal, and encoded according to an error-correcting code, comprising the steps of:

(a) receiving the audio signal;

(b) determining an initial position in the audio signal at which the message is initially detected;

(c) determining a bit error count for the message at each of a plurality of positions, including the initial position and at least one other position that is shifted relative to the initial position;

(d) determining an optimum position at which it is most probable that the message is embedded in the audio signal in accordance with the bit error counts; and

(e) recovering the message at the optimum position.

50. The method of claim 49, wherein:

the optimum position is determined according to the position that yields a bit error count that is a minimum among the plurality of positions.

51. The method of claim 49, wherein:

the optimum position is determined according to a mid-point of the positions that yield bit error counts that are minimums among the plurality of positions.

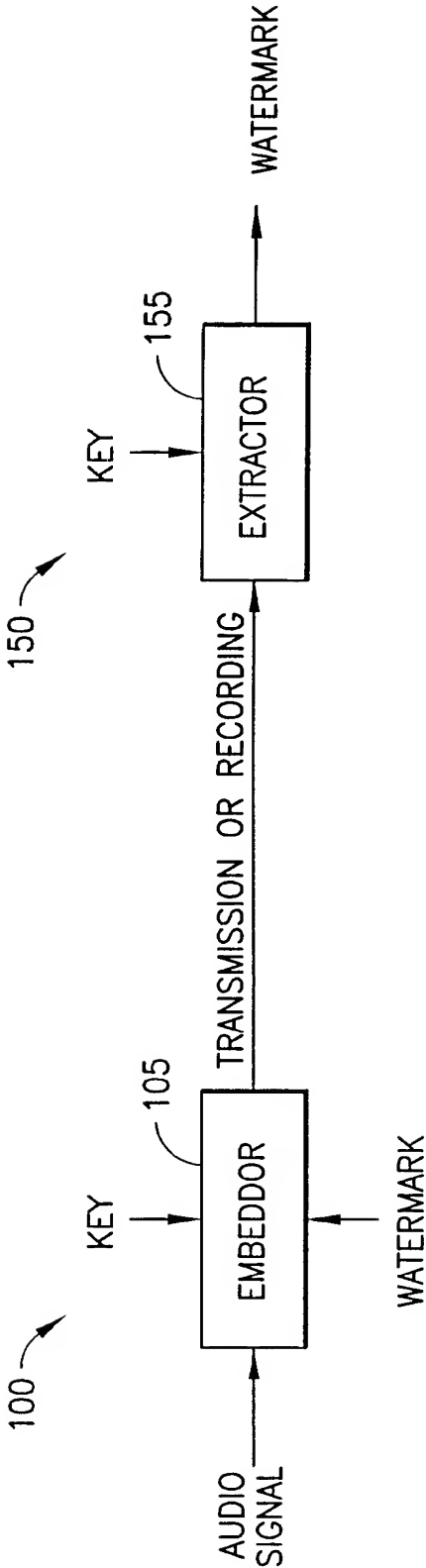


FIG.1

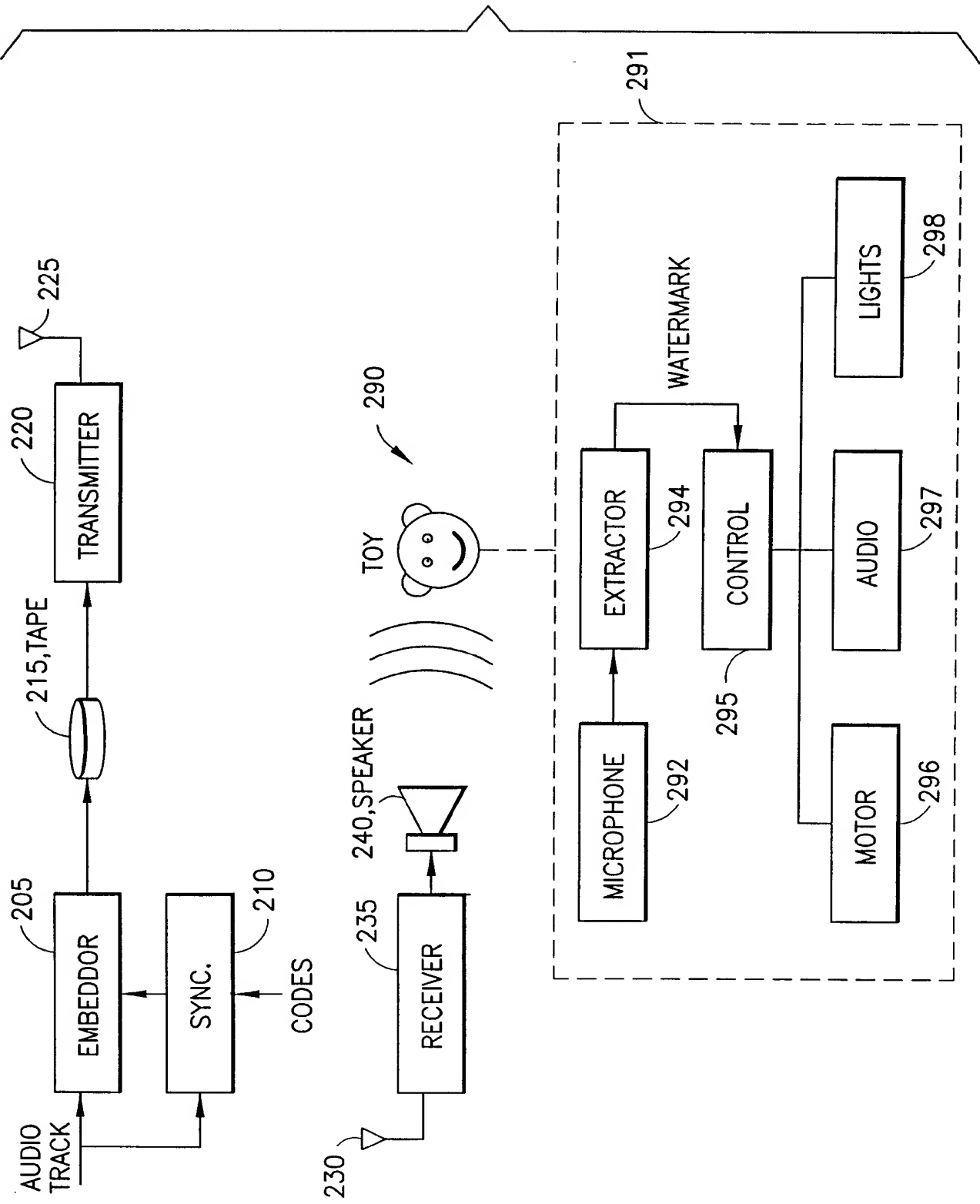
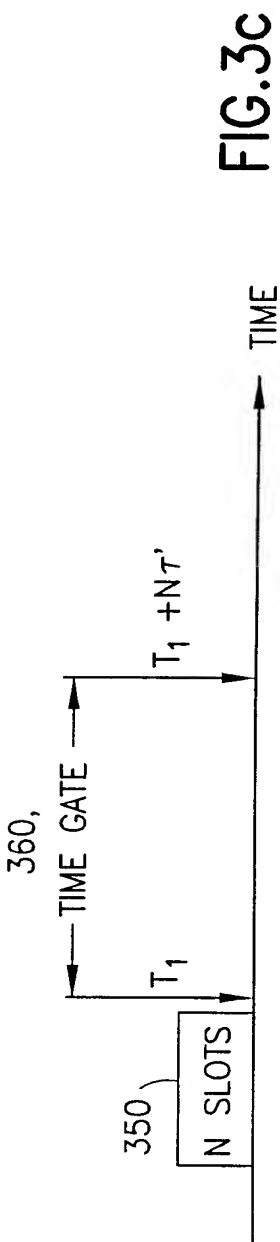
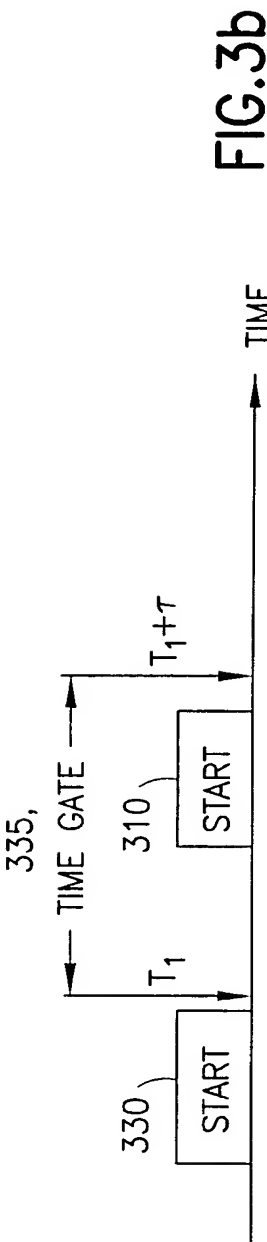
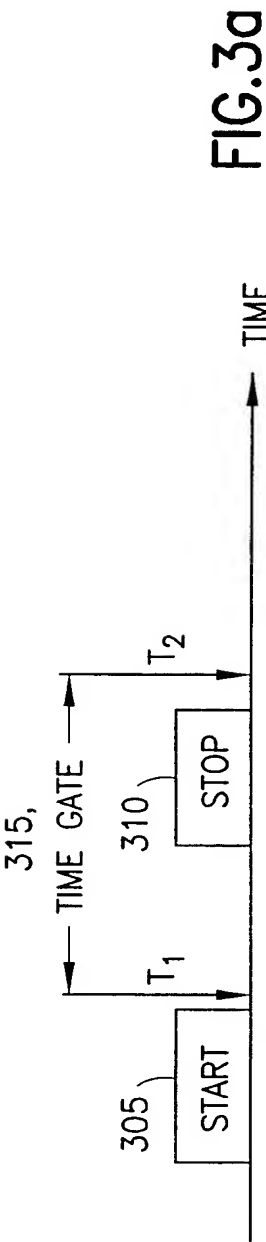


FIG.2



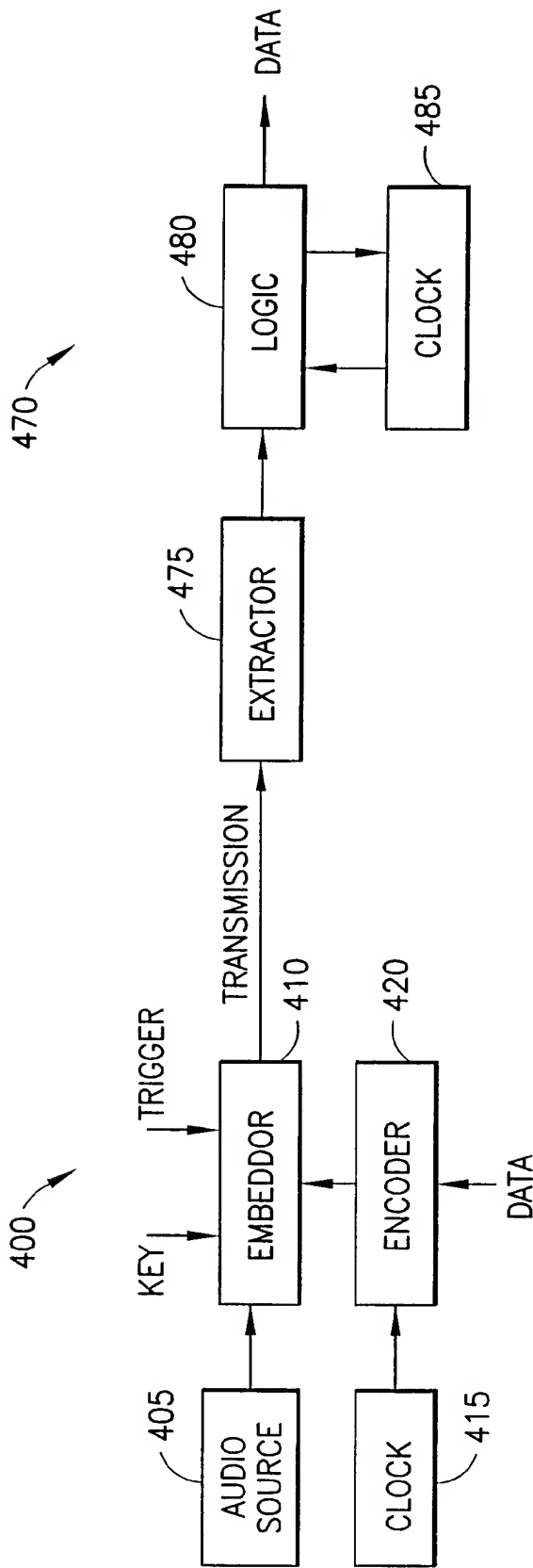


FIG.4

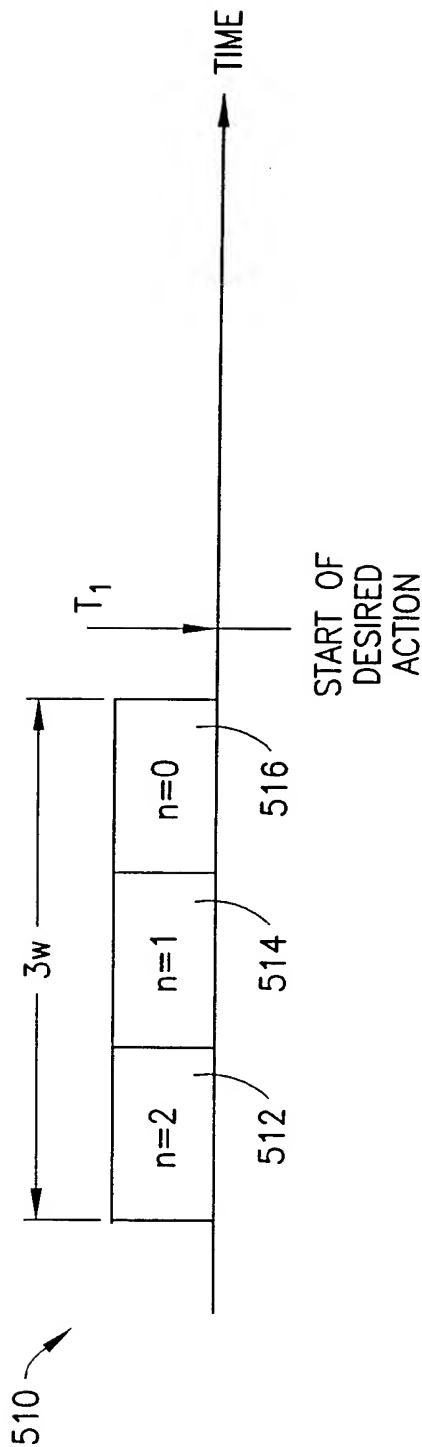


FIG.5

5/6

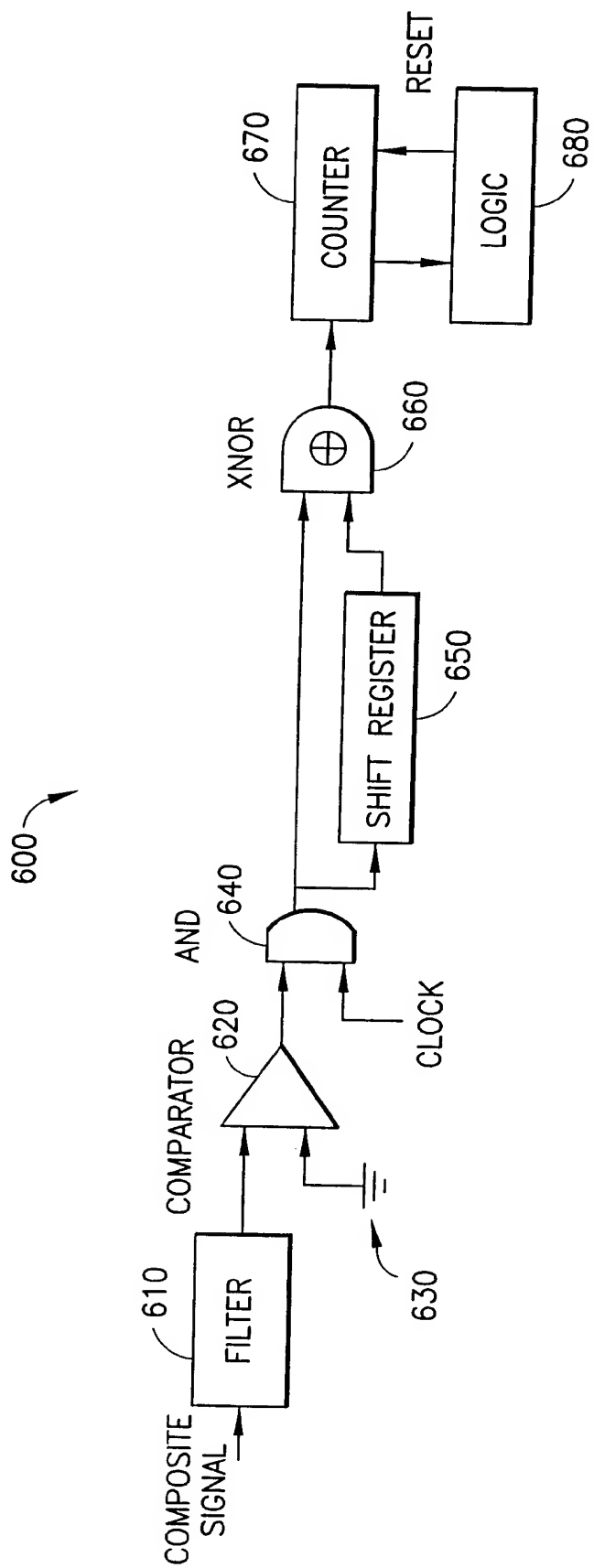


FIG.6

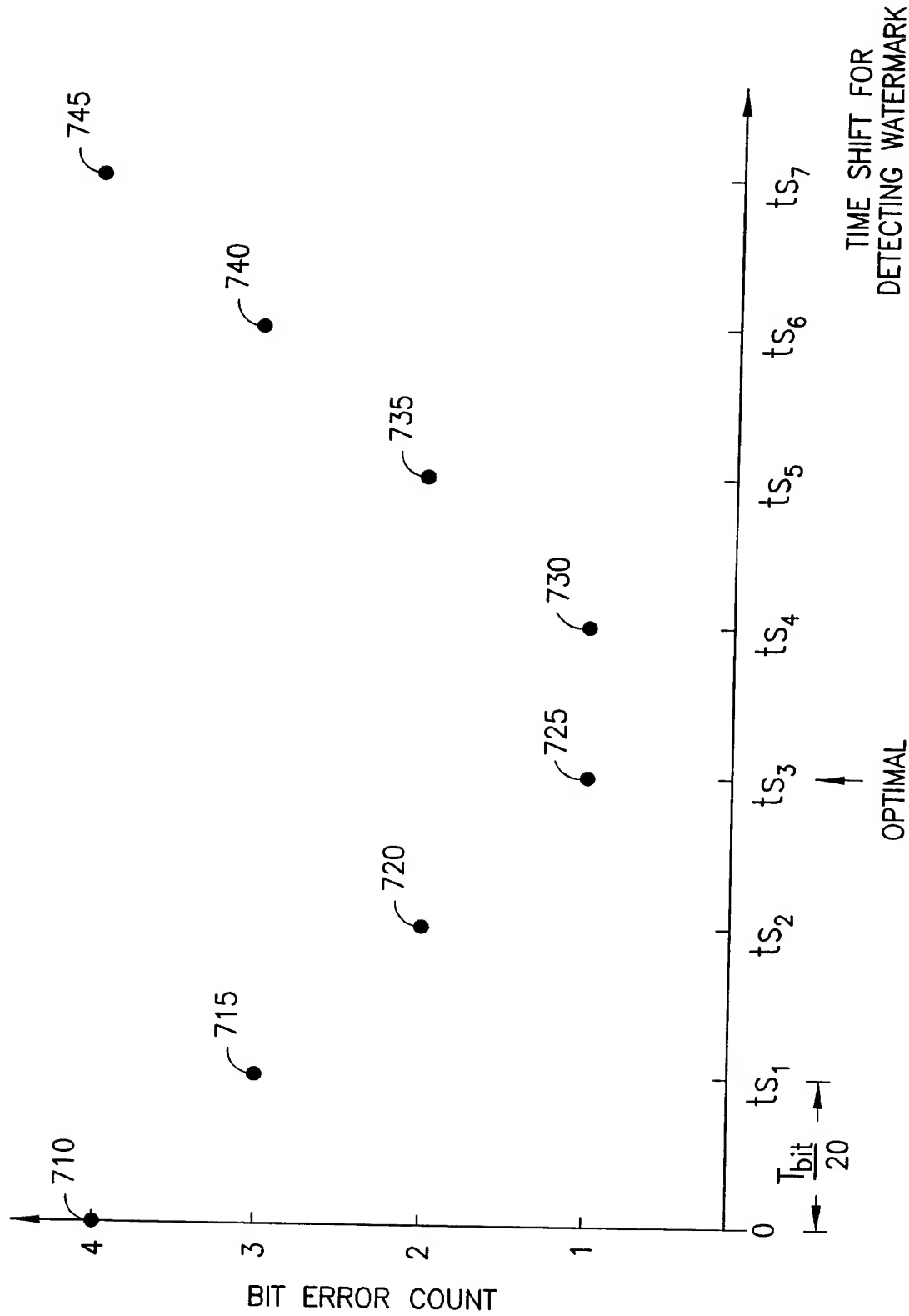


FIG.7